

Szabályalapú szintaktikai elemző szintaktikai szabályok nélkül

Kovács Viktória¹, Simkó Katalin Ilona², Szécsényi Tibor³

Szegedi Tudományegyetem, Bölcsészettudományi Kar

¹ viki921015@hotmail.com

² kata.simko@gmail.com

³ szecsényi@hung.u-szeged.hu

Kivonat: Cikkünkben bemutatjuk az általunk készített Prolog alapú magyar nyelvi szintaktikai elemzőt. A szabályalapú szintaktikai elemzéshez szükséges, nagyméretű lexikon kiküszöbölésére a magyarlánc morfológiai elemzőjét használtuk. A szabályrendszer leegyszerűsítésére öt alapszabályra vezetjük vissza az elemzést. Célunk, hogy megmutassuk, hogy szabályalapú elemzővel, kevés szabállyal is kezelni tudjuk a magyar nyelv sok sajátosságát.

1. Bevezetés

A természetes nyelvi kifejezések szintaktikai elemzését alapvetően kétféleképpen végezhetjük el. Statisztikai módszerrel egy nagy nyelvi korpusz alapján megjósolhatjuk, hogy egy adott mondatnak mi a legvalószínűbb szintaktikai szerkezete. Ilyenkor a korpusz mérete és változatossága garantálja, hogy a célmondatban meglevő minden konstrukciót felismerjen az elemző. Szabályalapú elemzőt használva olyan expliciten megfogalmazott helyettesítési szabályokra támaszkodhatunk, amiket az emberi nyelvtudás leírásánál felmerülő elméleti szempontok figyelembevételével alkottunk meg. Ezek a szabályok nem (csak) a megfigyelhető jelenségeket, mondatokat írják le, hanem a lehetőségeket.

A számítógéppel összegyűjtött, valószínűségi jellegű szabályok is egyfajta újraíró szabályoknak tekinthetők, azonban ezek a szabályok sem önmagukban, sem összességükben nem adnak pontos képet, magyarázatot az emberi nyelvek jellemzőiről, csak annak a visszaadására képesek, amiknek létrejöttüket köszönhetik: egy mondatnak, vagy az ő szintaktikai szerkezetének a valószínűségét határozzák meg. Így ugyanúgy kis valószínűségi értéket rendelnek egy ritkán használt konstrukciót tartalmazó, de lehetséges mondatnak/mondatszerkezetnek, mint egy gyakorihoz hasonló, de nem lehetségesnek. Használatukkal a prototipikus mondatok prototipikus elemzéseit kaphatjuk meg.

A szabályalapú elemzők ezzel szemben az emberi szabályalkotás miatt nem képesek az adatok ugyanolyan széleskörű figyelembevételére, mint a számítógépes, statisztikai alapú elemzők. Az elméleti, nyelvész szakemberek leginkább csak az érdekes, a számukra érdekesnek tűnő jelenségekre koncentrálnak. Az egyedi szabálymegállapítások mozaikjából azonban nem feltétlenül jön létre egy koherens szabályrendszer a nyelvre vonatkozóan. A leírt jelenségek bővülésével a leíró szabályrend-

szer is bővül, ezeknek a szabályoknak az egymásra gyakorolt hatását viszont nem mindig veszik figyelembe. A szabályalapú elemzők egyik legnagyobb hibaforrása a szabályok nagy száma.

Előadásunkban egy olyan szintaktikai elemzőt mutatunk be, ami a valószínű mondatok/mondatszerkezetek helyett a lehetséges mondatokat/mondatszerkezeteket elemzi, ugyanakkor a lehető legkevesebb szintaktikai szabályt alkalmazza. Elemzőnk csupán öt szintaktikai szabályt használ, de ezzel a magyar mondatok jelentős részét képes elemezni, egyúttal értelmezni is, úgy, hogy közben figyelembe veszi a magyar nyelv diskurzuskonfiguracionalitását és szabad szórendűségét is.

Az elemző egy számítógépes szintaxis szeminárium eredménye, így nem volt és nem is lehetett célunk, hogy egy teljes, minden jelenséget hatékonyan leíró elemzőt hozzunk létre, hanem csak az, hogy egy ilyen kevés szabállyal dolgozó, szabályalapú elemző megvalósíthatóságát, számítógépes implementálhatóságát bemutassuk.

2. Az elemző felépítése

Az elemző megalkotásánál arra törekedtünk, hogy a lehető legkevesebb újraíró szabályt kelljen alkalmaznunk. Ennek a célnak az eléréséhez a természetes nyelveknek azt a tulajdonságát használtuk ki, hogy a nyelvi kifejezésekben az egy-egy összetevőben előforduló szó szerkezetek számát és tulajdonságát mindig az összetevő egyik szava, az összetevő feje határozza meg: a főnévi csoportot a főnév, a névutós szerkezetet a névutó, a mondatot pedig az ige. A különböző lexikai egységek lexikai jellemzésében felsorolásként szerepel, hogy hány és milyen elemekkel kombinálható/kombinálódó. A fejjel kombinálódó más összetevők pedig a fejhez és egymáshoz viszonyítva viszonylag rögzített pozíciókat foglalhatnak el, és az így kialakítható szerkezetek száma az elméleti nyelvészeti kutatások szerint igen kicsi.

2.1. Kötött szórendű összetevők

A magyar nyelv rögzített szórendű szerkezeteinek – a főnévi csoportnak, a névutói csoportnak, a határozói és a melléknévi csoportnak – a leírására a Jackendoff [0] által javasolt X-vonás elmélet három szabálya elegendő. E három szabályból kettő a lexikai fej vonzatait helyezi el a szerkezetben, a specifikálót és a komplementumokat. A komplementumok az X fejhez jobbról csatlakoznak, az így kialakuló közbenső összetevőhöz, az X'-höz balról járul a specifikáló, így adva meg a teljes XP frázist. A kötött szórendű angol mondatszerkezet fejéül szolgáló *gives* igén mutatjuk be a két szabály működését:

A *gives* lexikai leírásában az ő szintaktikai tulajdonságainak a felsorolásán kívül (egyes szám harmadik személyű, ragozott, jelen idejű ige: V[fin, present, 3sg]) jelöljük, hogy milyen specifikálót, azaz igék esetében milyen alanyt kíván maga mellé (spec:(NP[nom,3sg])), illetve hogy milyen komplementumokat (egy tárgyat és egy prepozíciós kifejezést: comp:(NP[acc], PP[to])). A **komplementumszabály** a fej *gives* elemből és két neki komplementumként megfelelő elemből állítja össze a V' kifejezést:

$$X[\text{spec}:\alpha, \text{comps}:\beta] \rightarrow X[\text{spec}:\alpha, \text{comp}:(C)\oplus\beta] \quad C \quad (1)$$

ahol \oplus a konkatenáció jele. Mivel előre nem rögzített, hogy egy lexikai elemnek hány komplementuma lehet, a szabály egyszerre csak eggyel kapcsolja össze a fejet, a komplementumlistán legelöl levővel (a szabályban a C változó jelöli), a komplementumlista maradékát (β) pedig továbbadja a létrejövő összetett kifejezés komplementumlistájának. Így a komplementumszabály kétszeri alkalmazásával a *gives a book to Mary* igei csoport szerkezete $[[\text{gives } [a \text{ book}]] \text{ } [to \text{ Mary}]]$ lesz, a kategóriája pedig $V[\text{fin, present, 3sg, spec}:(NP[\text{nom, 3sg}]), \text{comp}:(\)]$, ahol a komplementumlista egy üres lista. Látható, hogy a szabály bal oldalán levő elem szintaktikai kategóriája és egyéb szintaktikai tulajdonságai (X), valamint a specifikálólístájának az elemei ($\text{spec}:\alpha$) megegyeznek a fej hasonló tulajdonságaival.

A másik X -vonás szabály a **specifikálósabály**, ami a fejből és a komplementumokból összeállított közbülső összetevőhöz csatolja a fej specifikálóját:

$$X[\text{spec}:\alpha, \text{comp}:(\)] \rightarrow A \quad X[\text{spec}:(A)\oplus\alpha, \text{comp}:(\)] \quad (2)$$

A szabály alkalmazásával a fej lexikai leírásában szereplő specifikálólístaról egyével kapcsolódnak az összetevők a fejből és a komplementumokból álló közbülső összetevőhöz. A specifikálósabály ismételt alkalmazásával kaphatunk teljes frázisokat (maximális projekciókat), azaz NP főnévi csoportokat, PP névutós szerkezeteket stb., ezeknek mind a specifikálólístájuk, mind a komplementumlistájuk üres lista.

Az X -vonás elmélet harmadik szabályának az ismertetésére, az adjunktumszabályra itt nem térünk ki, megvalósítása hasonló az előbbiekhöz.

2.2. Mondatszerkezetek

A magyar mondat szerkezet az eddig említett szerkezetekkel szemben nem kizárólag az X -vonás szabályrendszer szerint épül fel, hanem az igéből és a komplementumaiból álló, a komplementumszabály segítségével létrehozott igei csoportot megelőzőve különböző funkcionális pozíciók találhatók, a Topikok és a Fókusz pozíciója. Az itt megjelenő elemek maguk is az igei fej bővítményei, de szintaktikai és szemantikai tulajdonságaikban szelektáltak, illetve jelentésükben módosultak. Elemzőnkben a Topik és a Fókusz argumentumok a szerkezet fejének, az igének a lexikai leírásából származóan a komplementumlistához hasonlóan egy topik- és egy (egyelemű) fókuszlistán kerülnek felsorolásra. A *Péter Marinak adott egy könyvet* mondat *ad* igéjének lexikai leírásában az *egy könyvet* kifejezésnek megfelelő egyelemű komplementumlistán kívül még szerepel egy (esetünkben) egyelemű topiklista és egy szintén egyelemű fókuszlista is (az *ad* igének nincs specifikálója): $V[\text{fin, present, indef, 3sg, spec}:(\), \text{comp}:(NP[\text{acc, indef}]), \text{topic}:(NP[\text{nom, 3sg}]), \text{focus}:(NP[\text{dat}])]$. A csak igei fejű szerkezetek, azaz mondatok esetében alkalmazható **fókuszszabály** (3) és **topikszabály** (4) a már ismertetett két szabályhoz hasonló felépítésű, csak nem a komplementumlista vagy a specifikálólísta elemeit fogyasztják, hanem a topik- és a fókuszlista elemeit:

$$V[\text{spec}:(\), \text{comp}:(\), \text{topic}:\alpha, \text{focus}:\beta] \rightarrow F \quad V[\text{spec}:(\), \text{comp}:(\), \text{topic}:\alpha, \text{focus}:(F)\oplus\beta] \quad (3)$$

$$\begin{aligned} &V[\text{spec:}\langle \rangle, \text{comp:}\langle \rangle, \text{topic:}\alpha, \text{focus:}\langle \rangle] \rightarrow \\ &T \quad V[\text{spec:}\langle \rangle, \text{comp:}\langle \rangle, \text{topic:}\langle T \rangle \oplus \alpha, \text{focus:}\langle \rangle] \end{aligned} \quad (4)$$

Látható, hogy a (4) topikszabályt csak abban az esetben lehet alkalmazni, amennyiben az ige fej fókuszlistája üres, vagyis a topik-összetevők a mondatban megelőzik a fókuszt.

A bemutatott szabályok alkalmazásával, egy minimális szintaxis segítségével a magyar mondatok szintaktikai elemzését el tudjuk végezni. A mondatok szerkezete ezáltal az elméleti nyelvészeti szempontok szerint kialakított É. Kiss-féle [0] mondat-szerkezeti modellnek egy egyszerűsített változata lesz.

A szintaktikai szabályok számának minimalizálásának a feltétele az volt, hogy a szabályokat ne konkrét terminális és nem terminális kategóriák felhasználásával adjuk meg, hanem csak néhány sematizált kategóriával. A sematizált szintaktikai szabályoknak az ára a szavaknak, vagyis a lexikai elemeknek a gazdagabb jellemzése. Ugyanakkor egy jól működő szintaktikai elemző esetében a szavaknak ezeket az egyedi tulajdonságait úgyis mindenképpen fel kell tüntetni a lexikonban.

2.3. A szintaktikai elemző megvalósítása

Az ismertetett szintaktikai implementációja prolog programozási nyelven történt, azonban nem a prolog saját DCG formalizmusa szerint. Ennek az oka az, hogy mivel a szabályok nagyon sematikusak, ezért szükséges a balrekurzió miatt az nem alkalmazható. Ehelyett a left corner recognizer Blackburn és Striegnitz [0] által kidolgozott prolog implementációját alkalmaztuk az általunk kidolgozott szintaktikai szabályokkal.

3. Lexikai elemek, lexikai szabályok

A kevés szintaktikai szabály érdekében részletesen kidolgozott lexikai leírásokat kellett adni a lexikai elemekhez. Nem csak a szintaktikai kategóriát kellett feltüntetnünk, hanem a lexikai elemek egyéb szintaktikai-morfológiai tulajdonságait is, továbbá a lexikai elemek kombinációs képességét is. Azonban a lexikonnak csak a szintaxis felől nézve kell nagy lenni, mivel a lexikai elemek különböző morfológiai és argumentumszerkezeti variánsai a bázis lexikai elemekből lexikai szabályok segítségével levezethetőek. A bázis lexikon méretét egy morfológiai előelemző felhasználásával tovább csökkentettük.

3.1. Morfológia

A nagyméretű lexikon létrehozásának elkerülésére az elemzőnk a magyarlánc [0] morfológiai elemző modulját is felhasználja.

Az elemezni kívánt mondatokat először ezzel a külső, magyarlánc modullal elemezzük. A magyarlánc MSD-kódokkal látja el a mondat szavait a szófájuknak és

egyéb morfológiai jegyeiknek megfelelően. Az MSD-kód rengeteg morfológiai információt tartalmaz; ezekből a mi elemzőnk egyelőre nem mindet használja fel.

Az MSD-kódokból kinyerjük a szófaji meghatározást; főnevek esetén a számot, esetet és birtokviszonyt jelölő részeket; melléknevek esetén a számot, esetet és fokot; igéknél az igemódot, időt, számot, személyt és a tárgy határozottságára vonatkozó információt; ezek mellett az MSD-kód alapján elemezzük a névmások és határozott, valamint határozatlan determinánsok típusait is.

Manuálisan hozzáadott morfológiai információ csak az argumentumszerkezettel rendelkező elemek, azaz leginkább az igék argumentumszerkezetére vonatkozó részekben van az elemzőnkben. Mivel a magyarul a szavak szótövezését is elvégzi, a szótövek alapján egy kivétellista segítségével meghatározzuk, hogy melyik argumentumszerkezetes szó milyen argumentumszerkezet-típusba tartozik, majd a típusának megfelelő argumentumszerkezeti leírást rendelünk hozzá. A *látja* szóról a magyarul azt az információt adja vissza, hogy a töve *lát*, az MSD-kódja pedig *Vmip3s---y*. A *lát* tövű *Vxxxxxxx* MSD-kódú szavakhoz *vt*r tranzitív igei típust rendelünk, majd az MSD-kód alapján meghatározzuk a szó egyedi szintaktikai és morfológiai tulajdonságait (kijelentő módú, finit, jelen idejű, határozott ragozású, 3sg egyeztetési jegyű), majd a *vt*r típus és a már meghatározott jegyek generáljuk az argumentumszerkezetét: *comp:⟨NP[nom, 3sg], NP[acc, def]⟩*. Igék esetében a *spec*, *topic* és *focus* listák az alap lexikai leírásban mind üresek, azokat majd a később ismertetett lexikai szabályok töltik fel.

A kivétellistán nem szereplő, argumentumszerkezet nélküli vagy prototipikus argumentumszerkezettel rendelkező szavak lexikai leírását (főnevek, melléknevek, kötőszavak stb.) a szótő figyelembevétele nélkül, csak a szó MSD-kódjára támaszkodva hozzuk létre.

Az MSD-ből kinyert morfológiai információkat és az X-vonás elmélet alapszabályait összeillesztő egyszerű szabályok megalkotásával már képesek vagyunk egyszerű mondatok elemzésére intranszítív (1. példamondat), tranzitív (2. példamondat), ditranzítív igék (3. példamondat) és a kopula (4. példamondat) esetén is. Ezek a szabályok még kötött SVO szórenden működnek, ezeket bővítettük ki a nyelvspecifikus jelenségeket leíró szabályokkal.

1. A kutya fut.
2. Mari kergeti Pétert.
3. Mari kutyát ad Péternek.
4. A kutyák voltak pirosak.

3.2. Pro-drop

A magyar nyelv egyik tulajdonsága a pro-drop. Mivel az ige morfológiailag rengeteg információ megjelenik az alanyra és a tárgyra nézve, így azokat nem mindig szükséges expliciten megjeleníteni a mondatban. Megjelenhet az alanyi és a tárgyi összetevő is (5. példamondat); eltűnhet csak az alany (6. példamondat) vagy csak a tárgy (7. példamondat), de akár mindkét névszói összetevő is (8. példamondat).

5. Én látom őt.

6. Látom őt.
7. Én látom.
8. Látom.

Elemzőnket kibővítettük ennek a jelenségnek a kezelésére. Az igei csoportokat létrehozó szabályokat átalakítottuk, hogy alanyi és/vagy tárgyi összetevő hiányában is létre tudják hozni a mondatot. Egy-egy opcionális lexikai szabály segítségével a már létező igei lexikai elemek komplementumlistájáról törölhetjük az alanyi és a tárgyi főnévi csoportokat:

$$V[\text{comp}:\alpha \oplus \langle \text{NP}[\text{nom}] \rangle \oplus \beta] \Rightarrow V[\text{comp}:\alpha \oplus \beta] \quad (5)$$

Így a *látja* lexikai elemből, aminek a komplementumlistája $\text{comp}:\langle \text{NP}[\text{nom}, 3\text{sg}], \text{NP}[\text{acc}, \text{def}] \rangle$, a lexikai szabály $\text{comp}:\langle \text{NP}[\text{acc}, \text{def}] \rangle$ komplementumlistájú *levezetett* lexikai elemet képez.

3.3. Kopula

A magyarra jellemző, hogy a kopula harmadik személyű alany mellett jelen időben és kijelentő módban nem jelenik meg a mondat felszíni szerkezetében (9. példamondat), de egyéb esetben megjelenik (10. példamondat).

9. Ő katona.
10. Ő katona volt.

Ige nélküli mondatokban az elemzőnk képes alany esetű főnévből ige nélkül igei csoportot képezni. A főnév megtartja a komplementumlistáját, és ehhez hozzáadjuk az igei csoporthoz szükséges alanyesetű főnevet, vagyis az alanyt. Az így kapott igei csoport jelen idejű és harmadik személyű, számában pedig a főnévi predikátummal egyezik:

$$N[\text{nom}, \text{NUM}, \text{comp}:\alpha] \Rightarrow V[\text{fin}, \text{present}, 3\text{NUM}, \text{comp}:\langle \text{NP}[\text{nom}, \text{NUM}] \rangle \oplus \alpha] \quad (6)$$

Ilyen módon az *Ő katona*, ige nélküli mondatban a *katona* főnévből ($N[\text{nom}, 3\text{sg}, \text{comp}:\langle \rangle]$) egyes számú igei csoportot ($V[\text{fin}, \text{present}, 3\text{sg}, \text{comp}:\langle \text{NP}[\text{nom}, 3\text{sg}] \rangle]$) képezhetünk. Ennek a komplementumlistájára kerül fel az alany, ami az *ő* lesz a további mondat szerkezetben.

3.4. Topik és fókusz

A korábban bevezetett topik és fókusz szintaktikai szabály működését egy-egy lexikai szabály egészíti ki. Ezek a szabályok alakítják át az igeek eredeti argumentumszerkezetét úgy, hogy a megfelelő elemek a topik és fókusz pozíciókba kerülhessenek.

A (7) fókusz szabály egy ige üres fókuszlistájára rakhat át egy elemet a komplementumlistáról, ezzel biztosítva azt, hogy egy mondatban csak egy fókusz jelenjen meg az ige előtt. Az így a fókuszlistára mozgatott elemet fogja a szintaktikai fókusz szabály az ige előtt pozícióban elemezni.

$$V[\text{focus: } \langle \rangle, \text{comp: } \alpha \oplus \langle A \rangle \oplus \beta] \Rightarrow V[\text{focus: } \langle A \rangle, \text{comp: } \alpha \oplus \beta] \quad (7)$$

A (8) topik szabály nem csak üres topik listán működhet, egy mondatban több topik is megjelenhet az ige előtt. A szabály a listához kapcsol még egy elemet a komplementum listáról. A szintaktikai topik szabály ennek a listának az elemeit elemzi az ige előtt.

$$V[\text{topic: } \alpha, \text{comp: } \beta \oplus \langle A \rangle \oplus \gamma] \Rightarrow V[\text{topic: } \alpha \oplus \langle A \rangle, \text{comp: } \beta \oplus \gamma] \quad (8)$$

Ilyen módon a *Péter Marinak ad egy kutyát* mondat lehetséges elemzése a következő: az *ad* ige ($V[\text{comp: } \langle \text{NP}[\text{nom}, 3\text{sg}], \text{NP}[\text{acc}, \text{indef}], \text{NP}[\text{dat}]]$, $\text{topic: } \langle \rangle$, $\text{focus: } \langle \rangle$) komplementumlistájáról a lexikai fókusz szabály a *Marinak* összetevőnek megfelelő elemet a fókusz listára mozgatja ($V[\text{comp: } \langle \text{NP}[\text{nom}, 3\text{sg}], \text{NP}[\text{acc}, \text{indef}]]$, $\text{topic: } \langle \rangle$, $\text{focus: } \langle \text{NP}[\text{dat}] \rangle$), a lexikai topik szabály pedig a *Péter* összetevőt rakja át a topik listára ($V[\text{comp: } \langle \text{NP}[\text{acc}, \text{indef}]]$, $\text{topic: } \langle \text{NP}[\text{nom}, 3\text{sg}] \rangle$, $\text{focus: } \langle \text{NP}[\text{dat}] \rangle$). A komplementum listának így már csak egy eleme marad: az *egy kutyát*, ez az elem jelenik meg egyedül az ige után. A szintaktikai topik és fókusz szabályok a topik és fókusz listák tartalmát az ige előtt elemzik a megfelelő sorrendben.

3.5. Scrambling

Az ige utáni pozícióban az összetevők szabad sorrendben jelenhetnek meg a magyar mondatokban. Egy újabb lexikai szabállyal kezeltük ezt a jelenséget, ami biztosítja, hogy az itt lévő összetevők szabadon keveredhessenek.

11. Odaadta Mari Péternek a kutyát.
12. Odaadta Péternek Mari a kutyát.
13. Odaadta Mari a kutyát Péternek.

A scrambling szabály az igei csoportok komplementum listájának elemeit az összes lehetséges sorrendben előállítja, a lista elemeinek összes permutációját generálja:

$$V[\text{comp: } \alpha] \Rightarrow V[\text{comp: } \text{Perm}(\alpha)] \quad (9)$$

Ez a szabály hossza létre a magyarra jellemző ige utáni szabad szórendet. Így az *odaadta* ige komplementumlistáján szereplő három összetevőt: *Mari*, *Péternek*, *a kutyát* az összes lehetséges sorrendbe állítja az ige után, a három vonzat mind a hat lehetséges sorrendjével képes elemezni a mondatot.

3.6. A lexikai szabályok sorrendje

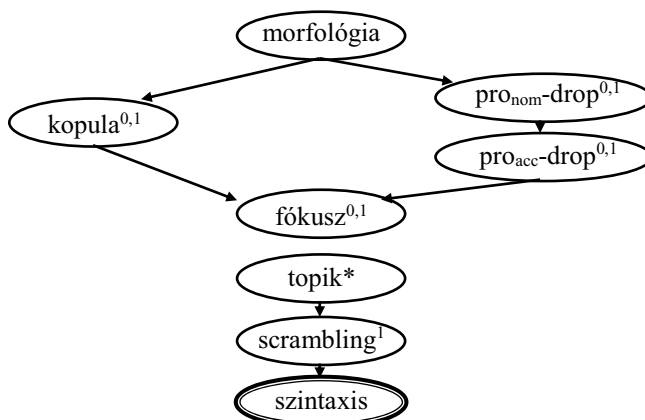
A 2. szakaszban megadott sematikus szintaktikai újraíró szabályok az így előkészített lexikai elemekből egyértelműen képesek felépíteni a mondatot (vagy képesek visszavezetni egy mondatot lexikai elemekre). A lexikai elemek viszont a lexikai szabályok miatt nem egyértelműek. Ugyanaz a lexikai elem többszörösen is elérhetővé válik a szintaktikai elemzésben, ha ugyanazon alap lexikai egységből különböző módon generáljuk. A *Péternek Mari ad egy kutyát* mondat *ad* igéjét például ugyanabból az *ad* alap lexikai egységből előállíthatjuk úgy, hogy először a (7) fókusz lexikai szabályt

alkalmazzuk, majd utána a (8) topik lexikai szabályt, vagy fordítva, először a topikot emeljük át a topiklistára, és csak utána a fókusz.

Az ilyen jellegű üres lexikai többértelműséget a lexikai szabályok hierarchiába rendezésével lehet megszüntetni: véges automataként szabályozzuk, hogy egy lexikai szabályt hányszor lehet alkalmazni, két lexikai szabályt milyen sorrendben kell elvégezni, illetve egy lexikai szabály alkalmazása milyen más szabályok alkalmazhatóságát zárja ki: a *zéró kopula* lexikai szabály nem alkalmazható az *alanyi* és a *tárgyi pro-drop* szabállyal együtt. Ha az *alanyi* és a *tárgyi pro-drop* lexikai szabályokat is alkalmazzuk, akkor az *alanyi pro-drop* szabálynak meg kell előznie a *tárgyi pro-drop* lexikai szabály alkalmazását. Ezek a lexikai szabályok csak a *fókusz* és a *topik* lexikai szabályok előtt alkalmazhatóak, a *fókusz* és a *topik* lexikai szabályok együttes alkalmazása esetén a *fókusz* szabály megelőzi a *topik* szabályt. A *scrambling* lexikai szabályt legutoljára kell alkalmazni.

A *scrambling* szabályt pontosan egyszer kötelező alkalmazni, a *topik* szabályt akárhányszor, az összes többi legfeljebb egyszer.

A lexikai szabályok sematikus rendezése (^{0,1}: legfeljebb egyszer; ¹: pontosan egyszer; *:akárhányszor alkalmazható):



4. Összegzés

Cikkünkben bemutattuk Prolog alapú magyar szintaktikai elemzőnk működését. Az elemző összesen öt szintaktikai szabállyal elemzi a mondatokat külső lexikon, a magyarlánc morfológiai elemző felhasználásával, így elkerülve a nagyméretű szótár használatát.

A jövőben tervezzük, hogy az elemzőt bővítjük. Terveink között szerepel az igék argumentumszerkezeteinek automatikus hozzáadása az elemzőhöz és összetett mondatok kezelése.

Hivatkozások

1. Blackburn, P., Striegnitz, K.: Natural Language Processing Techniques in Prolog. <http://cs.union.edu/~striegnk/courses/nlp-with-prolog/html/index.html> (2002)
2. É. Kiss, K: The Syntax of Hungarian. Cmabridge, Cambridge University Press (2002)
3. Jackendoff, R.: X-bar-Syntax: A Study of Phrase Structure. Linguistic Inquiry Monograph 2. Cambridge, MA: MIT Press (1977)
4. Zsibrita, J., Vincze, V., Farkas, R.: magyarlanc 2.0: szintaktikai elemzés és felgyorsított szófaji egyértelműsítés. In: Tanács, A., Vincze, V. (szerk.): IX. Magyar Számítógépes Nyelvészeti Konferencia. Szeged, Szegedi Tudományegyetem (2013) 368–374